The data is the sales volume from a retail store in recreation goods, clothing, footwear,

and chemist. (Extracted from https://timeseries.weebly.com/data-sets.html) The time range for this data set is the monthly sales from May 1995 to September 2010, a total of 185 data points for each of the four categories. Since the data selected are the sales records, considering the interest rate and inflation, we chose to build time series models upon the lognormal results of the data, which turns out to be the suitable fit compared to the original data.

Also, since the four categories of the products are sold in the same retail store, it is natural to assume that the sale of one category could affect another category. For instance, the total amount of money people spend in this area during a certain month could be limited, determined by the monthly income of local people, thus the over-sale of one category could result in the low sale of another category in that month or the following months. Another possible assumption is that the sale of the whole retail store could have a peak during certain period and higher than other months, like Black Friday, certain festivals, and promotions, which would increase the cash flow to all of the categories. Thus, we considered to use the Vector Auto Regressive (VAR) model to analyze the cross-relationship and lead-lag effect among the four categories, and to use Auto Regressive Integrated Moving Average (ARIMA) model to analyze the pattern of one category.